

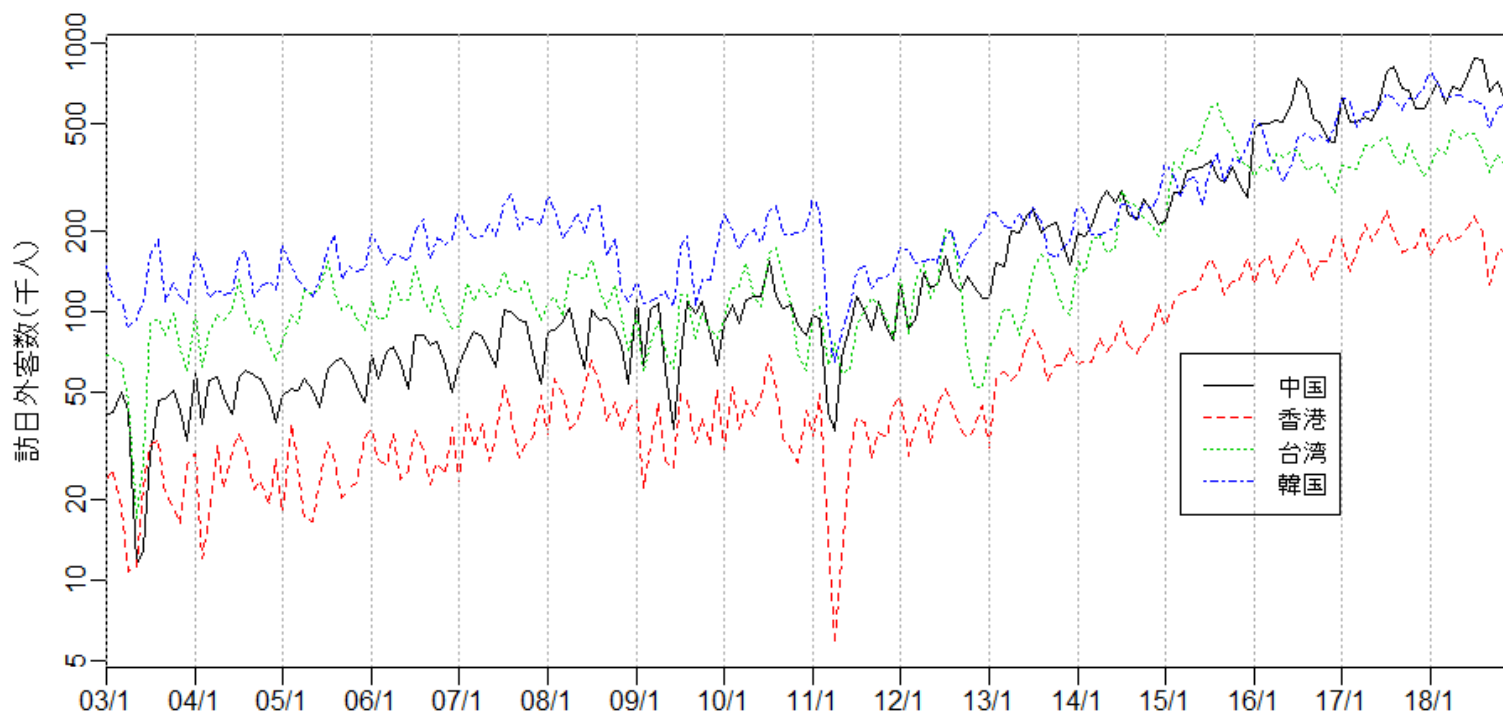
L-B1 課題説明

野村俊一
統計数理研究所

テーマ: 訪日外客数の分析と予測

- ◆ JNTO(日本政府観光局)より公開されている国別の訪日外客数の月次推移データについて、状態空間モデルを用いて増減分析 & 将来予測をしましょう
- ◆ データは下記URLから取得してください

<https://sites.google.com/site/shunichinomu/datkadai>



課題

- ◆ 訪日外客数の月次推移(国は自由選択)について、以下の分析を行い、レポート(形式自由)にまとめて提出してください
- 1. 基本構造時系列モデルを適用し、最も当てはまりの良いトレンド成分モデル＋季節成分モデルの組合せをAICで選択しましょう
- 2. 残差分析を行って異常値を検出し、その原因を考察した上で、対処しましょう
- 3. 上のいずれかのモデル、あるいは、他に考えられたモデルを用いて2019年1～12月の推移を予測し、予測値と95%予測区間を示しましょう(講評時に実際の数値を示します)

課題

4. 次の課題のうち2つ以上を選択して取り組みましょう
- ① カレンダー効果（曜日効果、閏年効果）を導入し、推定値および季節変動（季節成分＋カレンダー効果）の推移を示しましょう
 - ② 適当な2つの国のデータに対して、SUTSEモデルを適用し、観測値攪乱項および状態攪乱項の分散共分散行列の推定値を示しましょう
 - ③ 全ての国の訪日外客数の年次推移（別ファイル）に対して動的因子モデルをまとめて適用し、2つ以上の因子を抽出して因子の解釈をしましょう
 - ④ ポアソン分布を観測モデルとした非ガウス状態空間モデルを適用しましょう
 - ⑤ 上記以外に考えられるモデルの改良を試しましょう

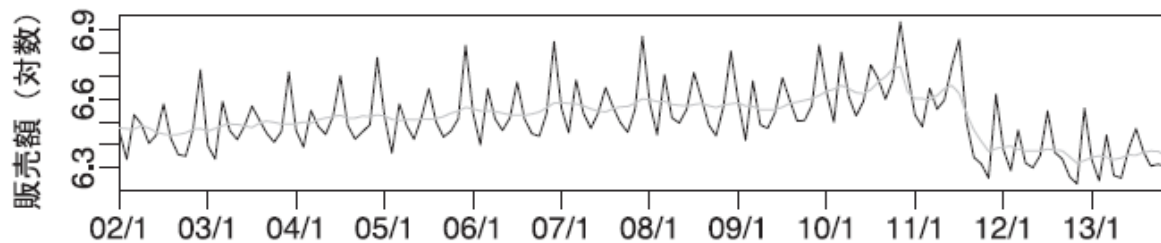
基本構造時系列モデルの解析例

- ◆ 織物衣服小売業の販売額データに、下表にある2通りずつのトレンド成分モデルと季節成分モデルの組合せによる基本構造時系列モデルをそれぞれ当てはめ、2011年までのデータから、2012～2013年を予測する
- ◆ AICによると、ローカルレベルモデル＋可変季節成分のモデルが最も当てはまりが良い結果となる

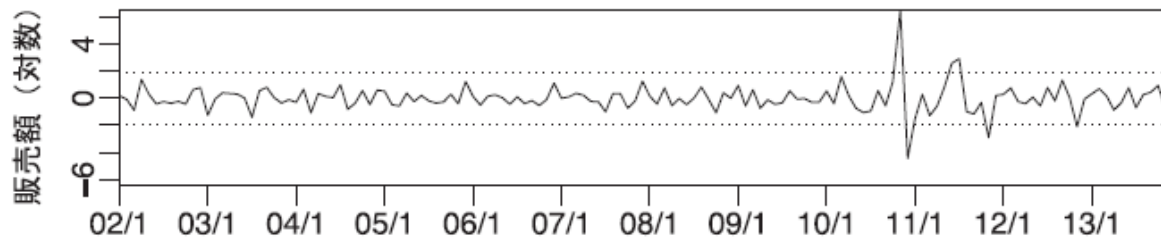
トレンド成分モデル	季節成分モデル	最大対数尤度	AIC	
ローカルレベルモデル	固定季節変動	-535.3	1098.7	
ローカルレベルモデル	可変季節変動	-531.6	1093.1	最良
平滑化トレンドモデル	固定季節変動	-535.4	1100.7	
平滑化トレンドモデル	可変季節変動	-533.3	1098.7	

状態空間モデルの残差分析の例

- ◆ 基本構造時系列モデルを当てはめた標準化残差を見ると、2010/11-12, 2011/6-11の残差が特に大きい
2010/12:エコポイントの制度変更 2011/7:地上アナログ放送停波

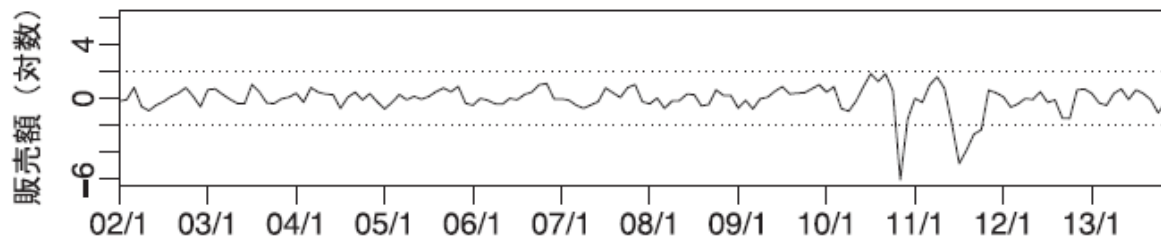


(a) 販売額の対数系列と平滑化状態の水準成分



(b) 観測値攪乱項の補助残差

標準正規分布
を仮定した
95%信頼区間
(-1.96, 1.96)



(c) 状態攪乱項 (水準成分) の補助残差

標準正規分布
を仮定した
95%信頼区間
(-1.96, 1.96)

異常な残差への対処

◆ 構造変化(干渉変数)の導入 + 異常値除外

$$y_t = \mu_t + \gamma_t + \beta x_t + \varepsilon_t$$

異常な残差に
対応する観測値
をNA(欠測)にする

$$x_t = \begin{cases} 1 & \text{2011年8月以降} \\ 0 & \text{2011年7月以前} \end{cases}$$

干渉変数
(水準シフト)

```
# 2010 年11 月, 12 月のデータを異常値として除外 (NAとして欠測値扱い)
machineNA <- machine
machineNA[sales[,1] %in% c("2010年11月", "2010年12月")] <- NA

# 2011 年8 月以降の水準シフト干渉変数の定義
ShiftLevel <- (1:nrow(sales) >= which(sales[,1]=="2011年8月"))
```

```
# 水準シフト干渉変数を加えたモデル
modShift <- SSMModel(log(machineNA) ~ SSMtrend(1, Q = NA)
  + SSMseasonal(12, Q = NA, sea.type="dummy" + ShiftLevel, H = NA)
fitShift <- fitSSM(modShift, numeric(3))
kfsShift <- KFS(fitShift$model,
  smoothing=c("state", "mean", "disturbance"))
```

ダミー変数
(0: 2011/7迄, 1: 2011/8以降)
を外来変数として加える

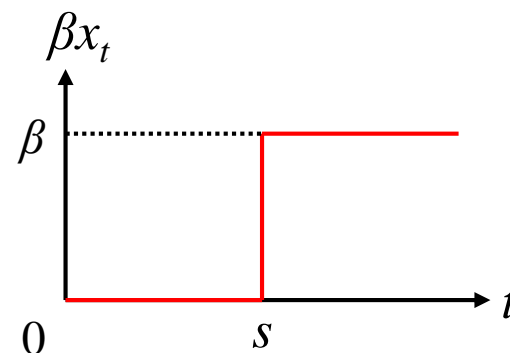
干渉変数

- ◆ 構造変化(干渉変数)のある構造時系列モデル

$$y_t = \mu_t + \gamma_t + \beta x_t + \varepsilon_t$$

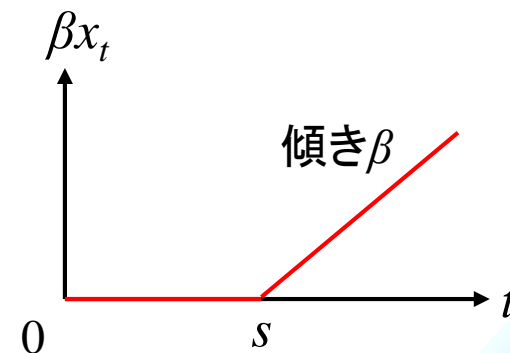
- ◆ 水準シフト干渉変数

$$x_t = \begin{cases} 1 & t \geq s \\ 0 & t < s \end{cases}$$



- ◆ 傾きシフト干渉変数

$$x_t = \begin{cases} t - s & t \geq s \\ 0 & t < s \end{cases}$$

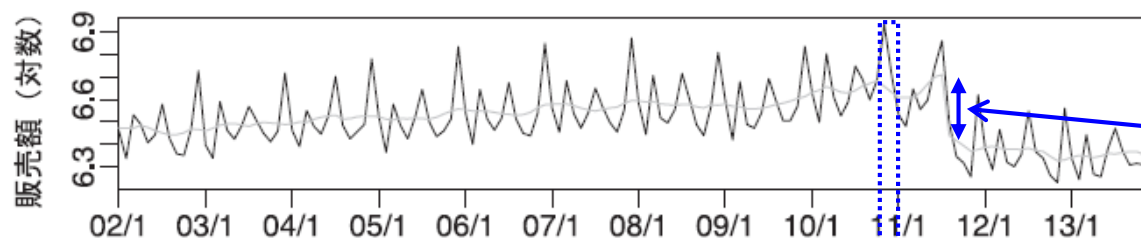


- ◆ 両者を併用してもよい

残差分析と構造変化(機械器具小売業)

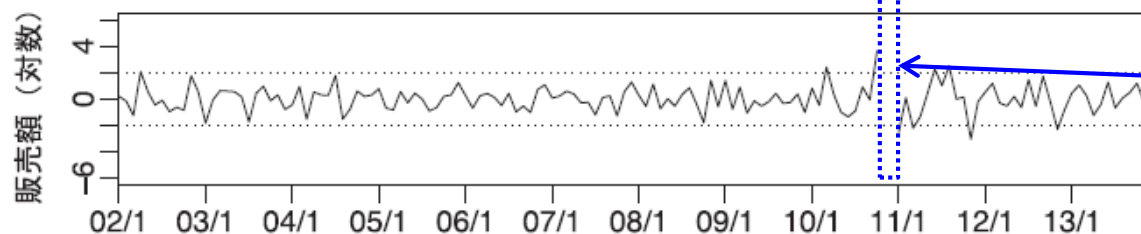
◆ 構造変化(干渉変数)の導入＋異常値除外

$$y_t = \mu_t + \gamma_t + \beta x_t + \varepsilon_t \quad x_t = \begin{cases} 1 & \text{2011年8月以降} \\ 0 & \text{2011年7月以前} \end{cases}$$



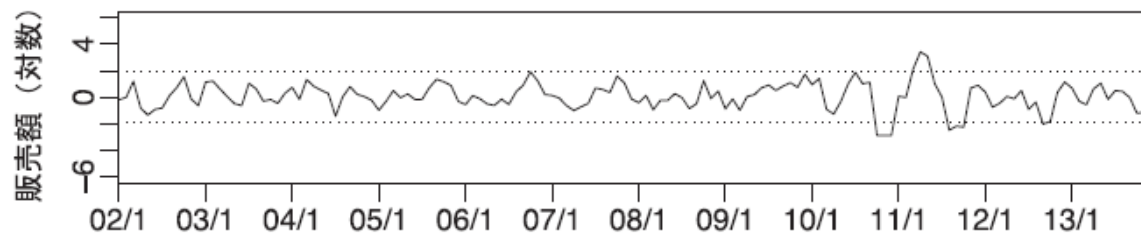
(a) 販売額の対数系列と平滑化状態の水準成分 (シフトを含む)

シフト幅 $\beta = -0.270$
($e^{-0.270} = 0.76$ 倍)



(b) 観測値攪乱項の補助残差

異常値を除外した期間。
観測値と補間値の差は
 $0.320 (e^{0.320} = 1.38$ 倍)と
 $-0.167 (e^{-0.167} = 0.85$ 倍)



(c) 状態攪乱項 (水準成分) の補助残差

長期予測

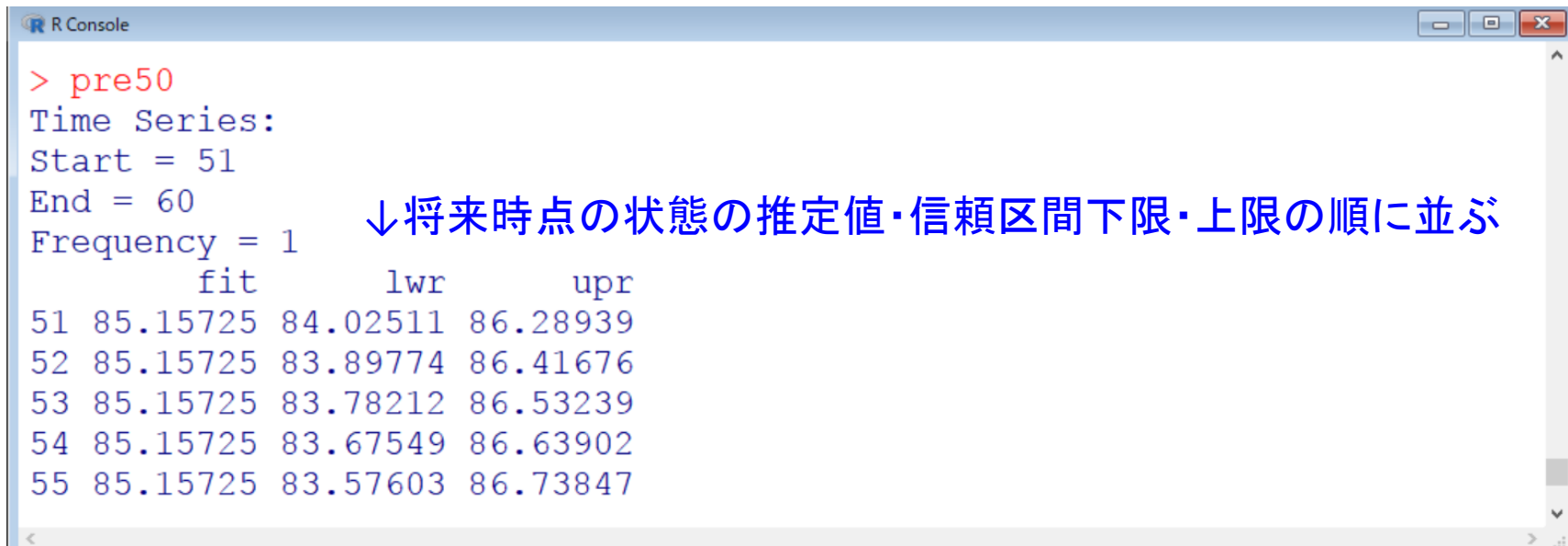
長期予測

```
mod50 <- SSMModel(Weight[1:50] ~ SSMtrend(1, Q = NA), H = NA)
fit50 <- fitSSM(mod50, numeric(2), method = "BFGS")
conf50 <- predict(fit50$model, interval = "confidence", n.ahead = 10)
pre50 <- predict(fit50$model, interval = "prediction", n.ahead = 10)
```

予測期間長

長期予測(予測期間を欠測値NAとして予測するやり方)

```
Weight50 <- Weight[c(1:50, rep(NA, 10))] # 51日目以降をNA(欠損)に置き換え
mod50NA <- SSMModel(Weight50 ~ SSMtrend(1, Q = NA), H = NA)
fit50NA <- fitSSM(mod50NA, numeric(2), method = "BFGS")
conf50NA <- predict(fit50NA$model, interval = "confidence", level = 0.95)
pre50NA <- predict(fit50NA$model, interval = "prediction", level = 0.95)
```



```
> pre50
Time Series:
Start = 51
End = 60
Frequency = 1
```

↓将来時点の状態の推定値・信頼区間下限・上限の順に並ぶ

	fit	lwr	upr
51	85.15725	84.02511	86.28939
52	85.15725	83.89774	86.41676
53	85.15725	83.78212	86.53239
54	85.15725	83.67549	86.63902
55	85.15725	83.57603	86.73847

カレンダー効果の解析例

◆ トренд成分＋季節成分＋閏年効果＋曜日効果

$$y_t = \mu_t + \gamma_t + \beta_{\text{閏}} x_{\text{閏}t} + \sum_{k=\text{月}, \dots, \text{土}} \beta_{k\text{曜}} (x_{k\text{曜}t} - x_{\text{日曜}t}) + \varepsilon_t$$

◆ 閏年効果と曜日効果（日曜を基準）の推定結果（10億円単位）

閏年	月	火	水	木	金	土	日
21.8	2.7	-12.1	1.7	-10.6	7.9	4.4	6.0

◆ 30日ある月の曜日効果（10億円単位）

月火	火水	水木	木金	金土	土日	日月
-9.4	-10.4	-8.9	-2.7	12.2	10.4	8.7

◆ 31日ある月の曜日効果（10億円単位）

月火水	火水木	水木金	木金土	金土日	土日月	日月火
-7.7	-21.0	-1.0	1.7	18.2	13.1	-3.4

月次死者数の解析: モデルの設計

◆ 指針に基づいて状態空間モデルを構築する

$$(1) \quad y_t \sim \text{Poisson}(\exp\{\mu_t + \gamma_t\})$$

$$(2) \quad \mu_t = \mu_{t-1} + \Delta\mu_t$$

$$(3) \quad \Delta\mu_t = \Delta\mu_{t-1} + \eta_{t1}, \quad \eta_{t1} \sim \text{Normal}(0, Q_1)$$

$$(4) \quad \gamma_t = -(\gamma_{t-1} + \dots + \gamma_{t-11}) + \eta_{t2}, \quad \eta_{t2} \sim \text{Normal}(0, Q_2)$$

観測モデル (1) : ポアソン分布

※ 対数リンク: $\log E(y_t | \mu_t, \gamma_t) = \mu_t + \gamma_t$

状態モデル $\begin{cases} (2), (3) : \text{平滑化トレンドモデル} \\ (4) : \text{季節成分モデル} \end{cases}$

月次死者数の解析: モデルの実装

◆ KFASならわずか4行で将来予測まで実行できる

月次死者数の解析コード

状態空間モデルの定義

```
modPois <- SSMModel(tsukiNA ~ SSMtrend(2, Q=list(0,NA))  
  + SSMseasonal(12, Q=NA), distribution="poisson")
```

ポアソン分布を指定

(観測値攪乱項分散Hの指定は不要となる)

最尤法による未知パラメータの推定

```
fitPois <- fitSSM(modPois, c(-15,-10), nsim=1000, method="BFGS")
```

いろいろな初期値を
試して大域的最適解
を探索する

非ガウスモデルの場合、
シミュレーション数を指定する
(ただしnsim=0でも近似解析できる)

状態推定

```
kfsPois <- KFS(fitPois$model, nsim=1000)
```

長期予測(予測区間)

```
prePoisNA <- predict(fitPoisNA$model, interval="prediction",  
  level=0.95, nsim=10000)
```

予測区間のシミュレーション数は多めにとる

2変量ローカルレベルモデルの実装と推定結果

↓体重と体脂肪率のデータを列結合

```
modSUTSE <- SSModel(cbind(Weight, Bodyfat) ~  
  SSMtrend(1, Q = matrix(NA,2,2)), H = matrix(NA,2,2))
```

Q,H は未知の 2×2 分散共分散行列

```
fitSUTSE <- fitSSM(modSUTSE, numeric(6), method="BFGS")
```

Q,H は対称行列で各々3つの未知成分をもつため、合計6つの未知パラメータ

```
kfsSUTSE <- KFS(fitSUTSE$model)
```

- ◆ 推定された分散共分散行列 Q , H は以下のとおり
- ◆ 計測誤差・増減ともに体重・体脂肪率の間に相関をもつ

$$H = \begin{pmatrix} 0.1429 & 0.0224 \\ 0.0224 & 0.1836 \end{pmatrix},$$

相関係数

0.138

$$Q = \begin{pmatrix} 0.0775 & 0.0402 \\ 0.0402 & 0.0447 \end{pmatrix}$$

相関係数

0.683

動的因子モデル(2因子ローカルレベルモデル)

- ◆ 因子を増やすことで、1950年前後の変化(若年層の大幅低下)と近年の変化(全年齢層での低下)の違いが明らかになった

